

<p>Nama: <b>Ibnu Fajar Setiawan</b></p> <p>NIM: <b>065002000006</b></p>	 <p><b>Praktikum Data Warehouse</b></p>	<h1>MODUL 10</h1> <p>Nama Dosen: <b>Ir. Teddy Siswanto, MMSi</b></p> <p>Nama Asisten Laboratorium:</p> <ul style="list-style-type: none"><li>1. Azhar Rizki Zulma <b>065001900001</b></li><li>2. Nadiya Amanda Rizkania <b>064001900003</b></li></ul>
---	--	---

## Proyek Akhir Praktikum II

### 1. Teori Singkat

Data warehouse adalah jenis sistem manajemen data yang dirancang untuk memungkinkan dan mendukung kegiatan business intelligence (BI), terutama analitik. Gudang data semata-mata dimaksudkan untuk melakukan kueri dan analisis dan sering berisi sejumlah besar data historis. Data dalam gudang data biasanya berasal dari berbagai sumber seperti file log aplikasi dan aplikasi transaksi. Gudang data memusatkan dan mengkonsolidasikan sejumlah besar data dari berbagai sumber. Kemampuan analitisnya memungkinkan organisasi untuk memperoleh wawasan bisnis yang berharga dari data mereka untuk meningkatkan pengambilan keputusan. Seiring waktu, ia membangun catatan sejarah yang dapat sangat berharga bagi para ilmuwan data dan analis bisnis. Karena kemampuan ini, gudang data dapat dianggap sebagai "sumber kebenaran tunggal" organisasi.

### Cleansing Data

Cleansing Data adalah suatu proses mendeteksi dan memperbaiki (atau menghapus) data set, tabel, dan database yang korup atau tidak akurat. Data tidak akurat tersebut berupa data yang kosong dan tidak memiliki nilai atau berisi nilai NULL/NaN. Tujuan dari cleansing data ialah agar ketika data diproses nanti, tidak terjadi adanya error dari pengolahan data yang dilakukan akibat data yang tidak lengkap ataupun data hilang (missing value).

Berikut ini adalah langkah-langkah dalam melakukan cleansing data menggunakan Spoon Pentaho:



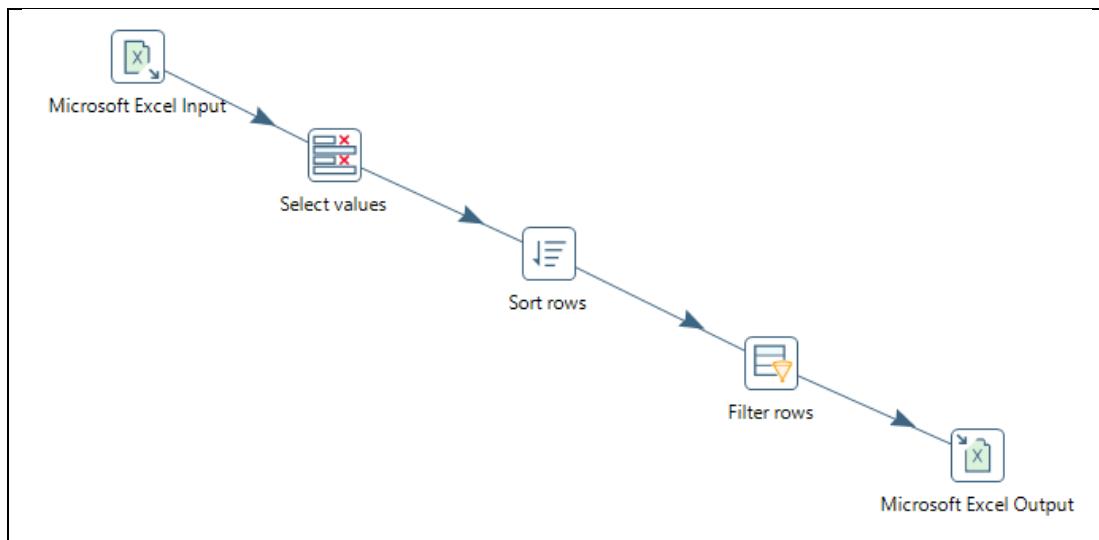
## 1. Data Preparation

Unduh serta ekstrak datanya dan satukan kedalam folder seperti pada gambar dibawah.

	Name	Date modified	Type	Size
ss	Cabai Merah Keriting Desember 2018	14/11/2020 21:16	Microsoft Excel C...	10 KB
ds	Cabai Merah Keriting November 2018	14/11/2020 21:14	Microsoft Excel C...	10 KB
nts	Cabai Merah Keriting Oktober 2018	14/11/2020 21:02	Microsoft Excel C...	10 KB
tp	Cabai Merah Keriting September 2018	14/11/2020 20:45	Microsoft Excel C...	10 KB
	data cabai keriting merah 2018 (jan ags)	01/05/2021 12:10	Microsoft Excel 97...	108 KB

## 2. Excel Input

- Buka Pentaho Spoon Data Integration lalu buat struktur seperti pada gambar dibawah ini lalu simpan dengan format ***Cleansing Cabai Merah Keriting [Bulan] [Tahun].ktr*** seperti ***Cleansing Cabai Merah Keriting Agustus 2019.ktr***



- Berikut ini adalah konfigurasi pada step *Microsoft Excel Input* (Files, Sheets, Fields)

The screenshot shows two windows of the Microsoft Excel Input configuration interface.

**Top Window (Files Tab):**

- Step name:** Microsoft Excel Input
- Spread sheet type (engine):** Excel 97-2003 XLS (JXL)
- Selected files:** C:\Users\Azhar Rizky Zulma\OneDrive\Documents\Job\Data Warehouse\Praktikum Video Cleansing and Filter
- Accept filenames from previous steps:**
- Accept filenames from previous step:**
- Step to read filenames from:** (dropdown menu)
- Field in the input to use as filename:** (dropdown menu)

**Bottom Window (Sheets Tab):**

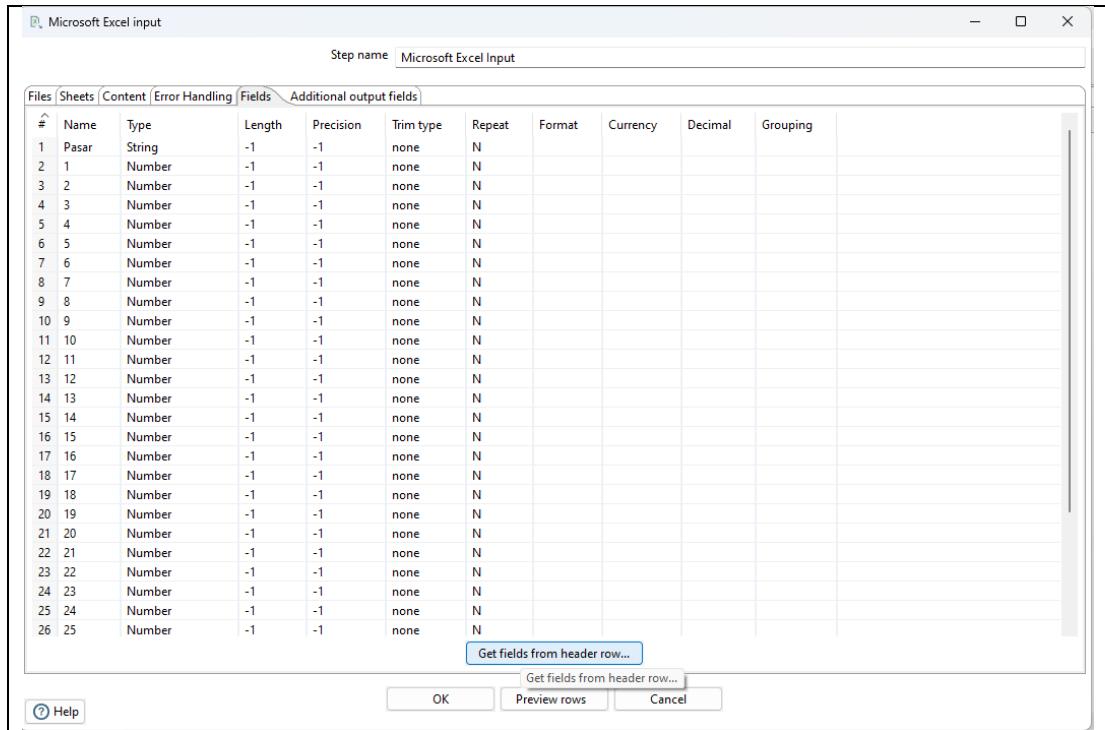
- Step name:** Microsoft Excel Input
- List of sheets to read:**

#	Sheet name	Start row	Start column
1	agustus 2018	0	0

- Enter list dialog:**

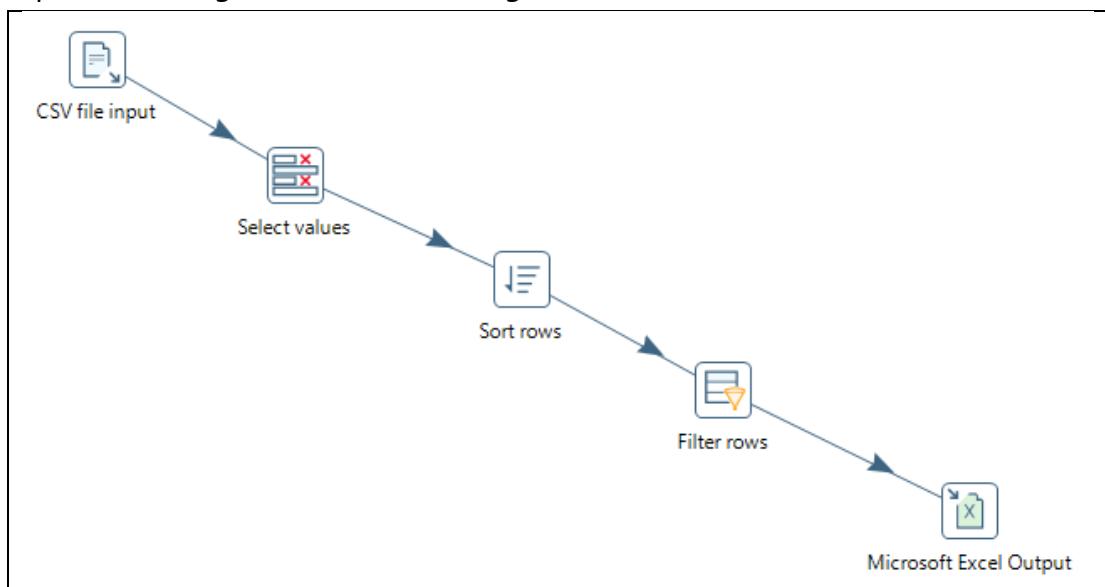
  - Available items:** januari 2018, Februari 2018, maret 2018, april 2018, mei 2018, juni 2018, juli 2018
  - Your selection:** agustus 2018
  - Buttons:** >, <, <<, OK, Cancel



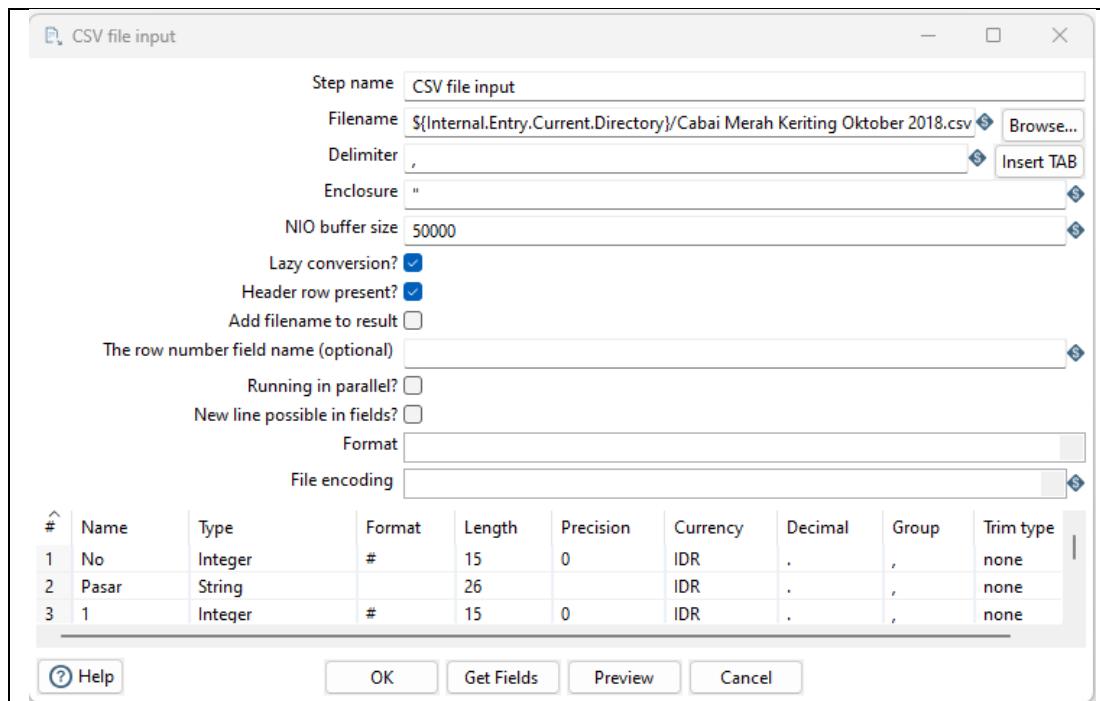


### 3. CSV Input

- Buka Pentaho Spoon Data Integration lalu buat struktur seperti pada gambar dibawah ini lalu simpan dengan format **Cleansing Cabai Merah Keriting [Bulan] [Tahun].ktr** seperti **Cleansing Cabai Merah Keriting Oktober 2019.ktr**

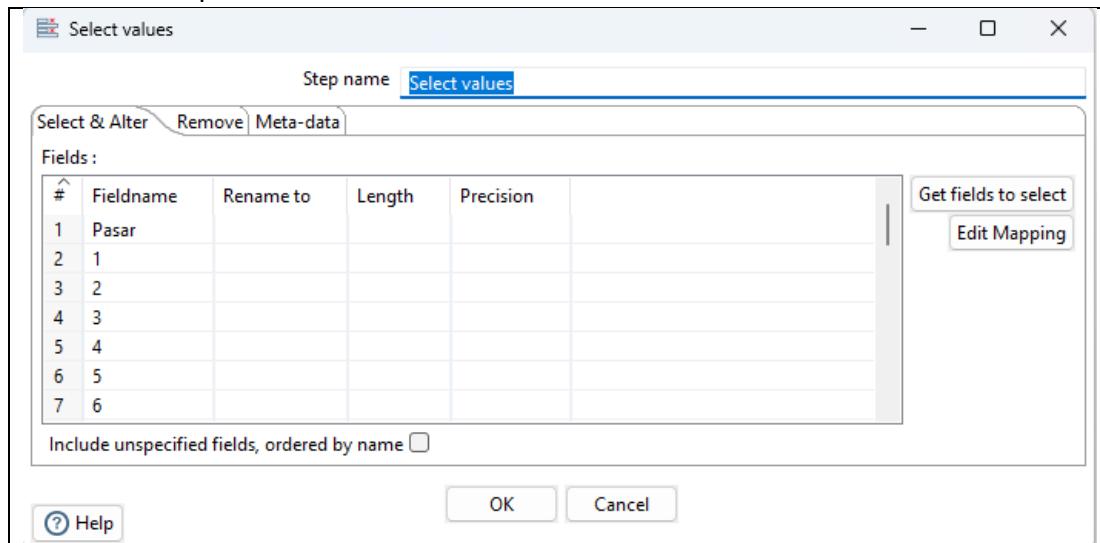


- Berikut ini adalah konfigurasi pada step CSV File Input.

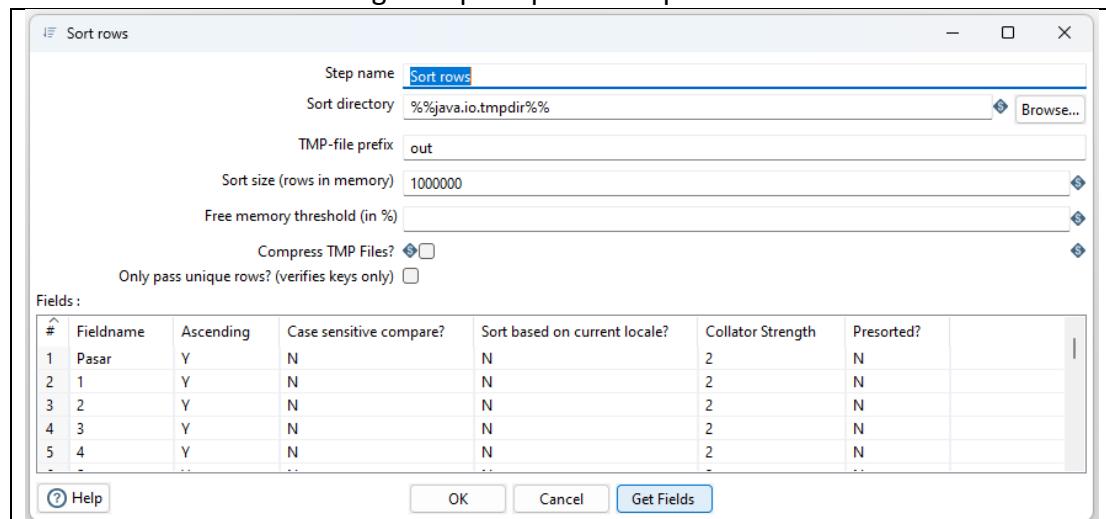


#### 4. Data Filtering

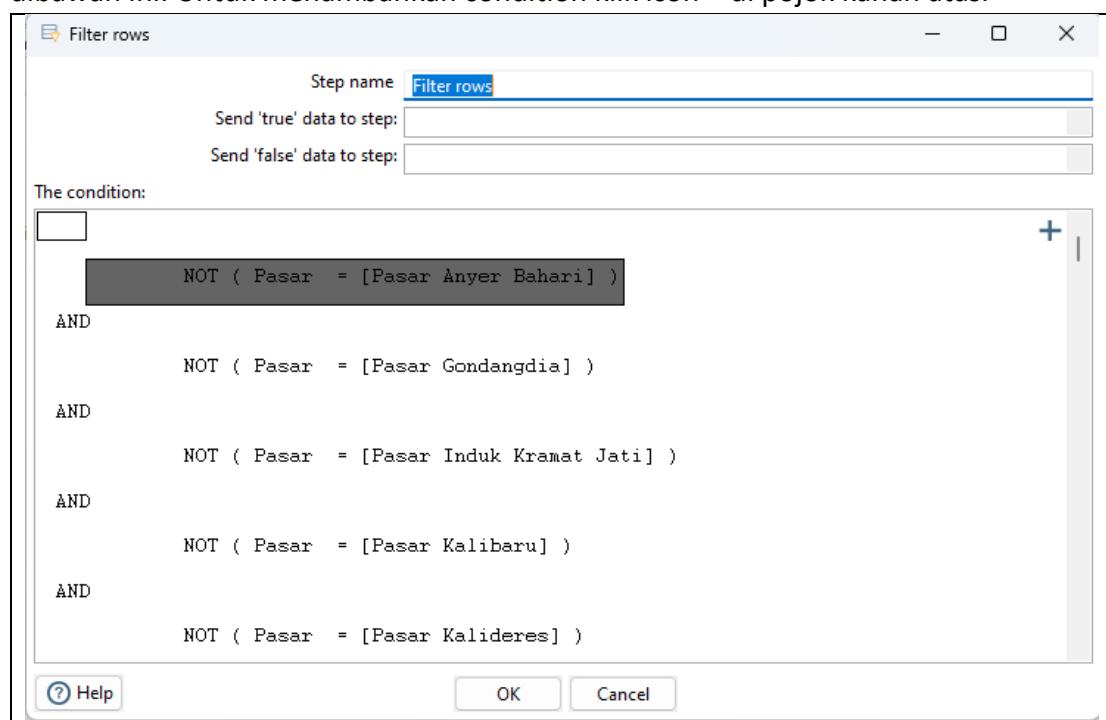
- Step selanjutnya adalah melakukan select value yang berguna untuk memilih kolom pada excel seperti pada gambar berikut. Kemudian Delete kolom No saja, karena nomor tidak diperlukan lalu Klik OK.

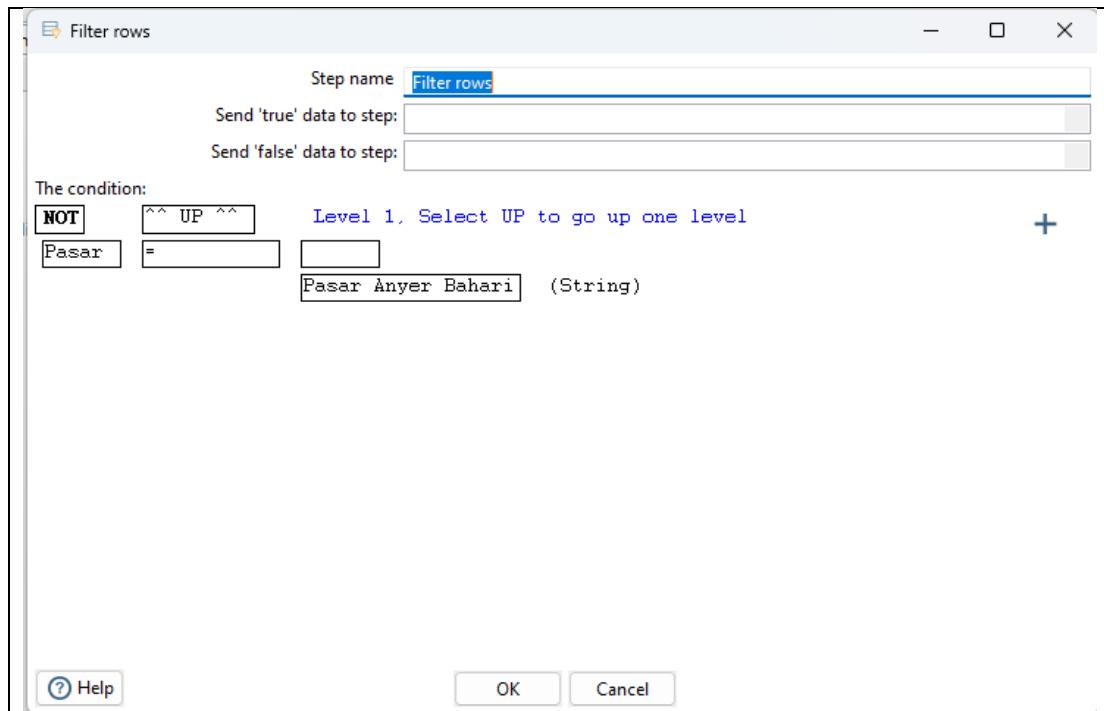


- Kemudian lakukan ascending data pada proses step Sort Rows



- Kemudian pada step Filter Rows, buang beberapa nama pasar seperti contoh pada dibawah ini. Untuk menambahkan condition klik icon + di pojok kanan atas.





- Berikut List Data Nama Pasar yang mau dibuang (2019)

	NOT ( Pasar = [Pasar Anyer Bahari] )
AND	NOT ( Pasar = [Pasar Gondangdia] )
AND	NOT ( Pasar = [Pasar Induk Kramat Jati] )
AND	NOT ( Pasar = [Pasar Kalibaru] )
AND	NOT ( Pasar = [Pasar Kalideres] )
AND	NOT ( Pasar = [Pasar Klender SS] )
AND	NOT ( Pasar = [Pasar Koja Baru] )
AND	NOT ( Pasar = [Pasar Rawamangun] )
AND	NOT ( Pasar = [Pasar Pramuka] )
AND	NOT ( Pasar = [Pasar Pluit] )
AND	NOT ( Pasar = [Pasar Pulo Gadung] )
AND	NOT ( Pasar = [Pasar Pesanggrahan] )
AND	NOT ( Pasar = [Pasar Glodok] )
AND	NOT ( Pasar = [Pasar Matraman KK] )
AND	NOT ( Pasar = [Pasar Jatinegara] )



## 5. Excel Output

Langkah terakhir simpan data menggunakan step *Microsoft Excel Output* dengan konfigurasi sebagai berikut lalu simpan dengan format nama **Cabai Merah Keriting [Bulan] [Tahun].xls** seperti **Cabai Merah Keriting Agustus 2019.xls** (File & Sheet)

The image shows two overlapping windows for configuring Microsoft Excel output. The top window is titled 'Microsoft Excel output' and has a 'Step name' of 'Microsoft Excel Output'. It includes fields for 'Filename' (C:\Users\Azhar Rizky Zulma\OneDrive\\$), 'Extension' (xls), and various options like 'Include stepnr in filename?' and 'Specify Date time format'. The bottom window is also titled 'Microsoft Excel output' with the same 'Step name'. It displays a table of fields with columns '#', 'Name', 'Type', and 'Format'. The fields listed are:

#	Name	Type	Format
1	Pasar	String	
2	1	Integer	
3	2	Integer	
4	3	Integer	
5	4	Integer	
6	5	Integer	



## 2. Alat dan Bahan

Hardware : Laptop/PC

Software : Spoon Pentaho from Hitachi Vantara

## 3. Elemen Kompetensi

a. Tugas Akhir Praktikum II – Screenshot Struktur yang sudah berhasil berjalan

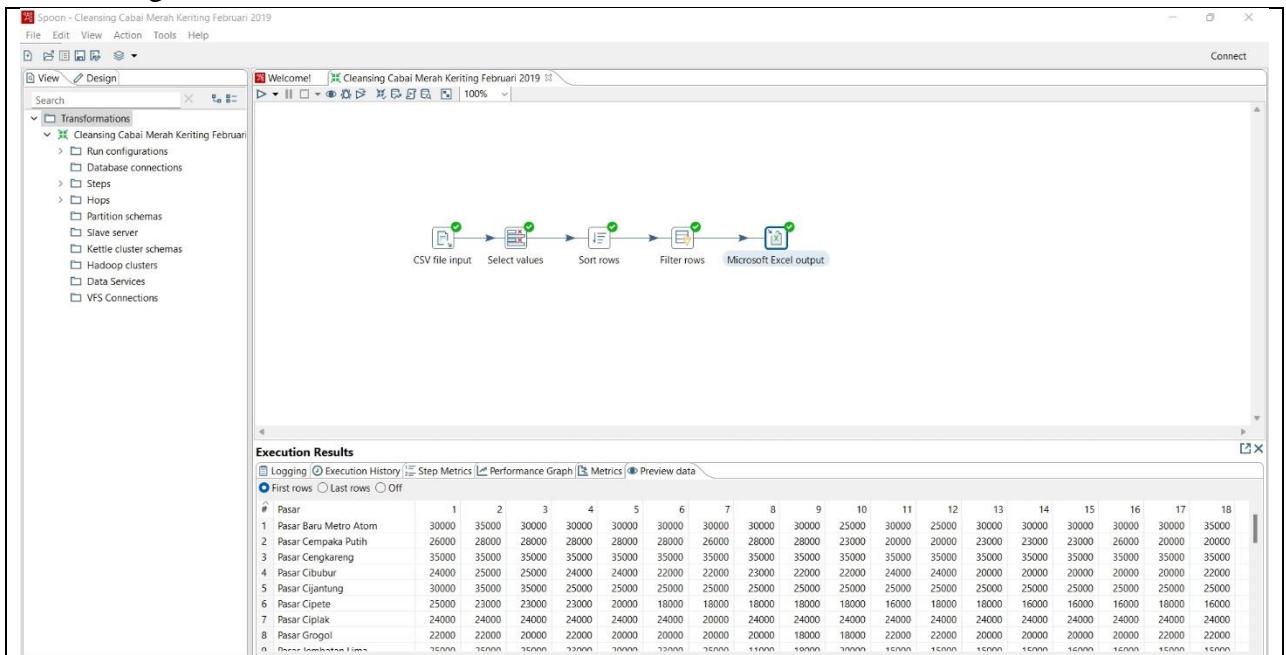
### 1. Cleansing Januari 2019

The screenshot displays the Spoon - Cleansing Cabai Merah Keriting Januari 2019 interface. The main window shows a workflow diagram with five steps: CSV file input, Select values, Sort rows, Filter rows, and Microsoft Excel output. The 'Execution Results' pane shows the first few rows of data, which is then displayed in an Excel spreadsheet titled 'Cabai Merah Keriting Januari 2019.xlsx'.

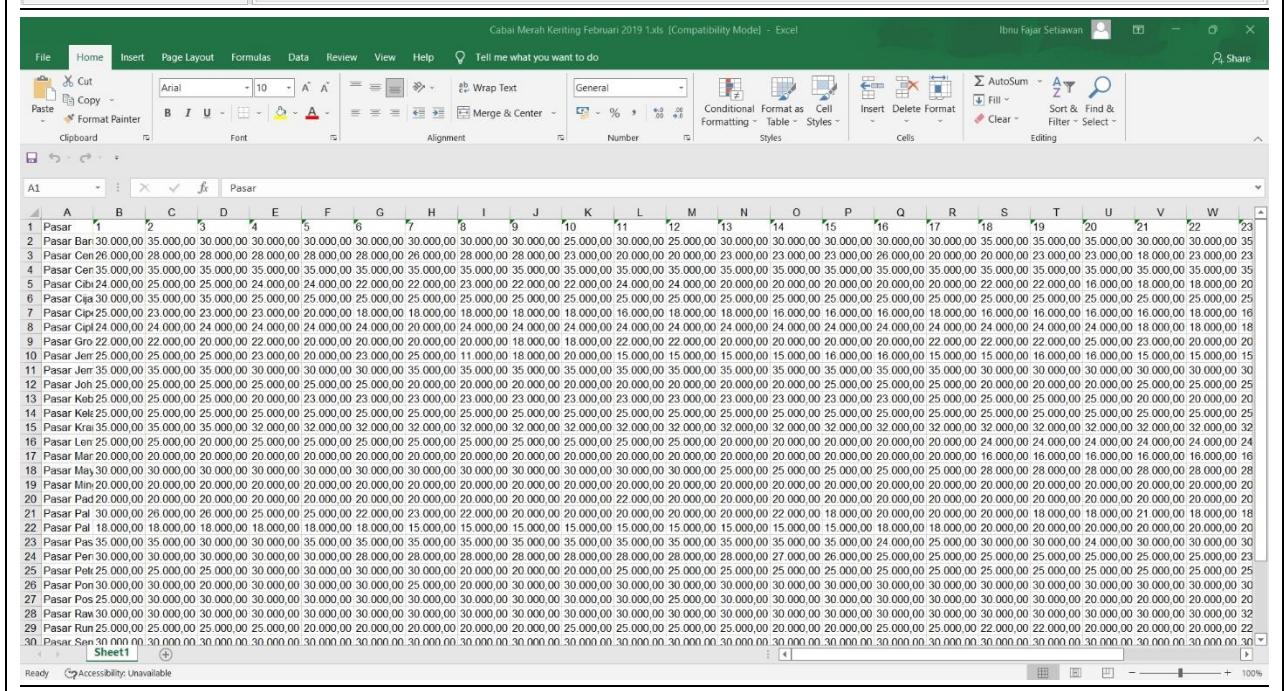
#	Pasar	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	Pasar																
2	Pasar Baru Metro Atom	45000	50000	50000	40000	50000	45000	45000	40000	40000	45000	45000	40000	40000	40000	40000	40000
3	Pasar Cempaka Putih	30000	30000	35000	35000	35000	35000	35000	30000	30000	35000	32000	30000	35000	35000	35000	35000
4	Pasar Cengkareng	30000	30000	30000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000
5	Pasar Cibitung	30000	30000	34000	34000	28000	28000	28000	28000	28000	28000	28000	28000	28000	28000	28000	28000
6	Pasar Ciputat	35000	35000	35000	37000	36000	36000	35000	32000	32000	32000	32000	32000	32000	32000	32000	32000
7	Pasar Cipatik	30000	30000	30000	30000	24000	24000	30000	30000	30000	28000	28000	28000	30000	30000	30000	30000
8	Pasar Grogol	30000	30000	30000	30000	30000	30000	30000	30000	30000	28000	28000	28000	25000	25000	25000	25000
9	Pasar Tambakrejo	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000



## 2. Cleansing Februari 2019



The screenshot shows the Kettle Data Integration tool interface. On the left, the 'Transformations' tree view is expanded to show the 'Cleansing Cabai Merah Keriting Februari' project. A process flow diagram is displayed, consisting of five steps: 'CSV file input', 'Select values', 'Sort rows', 'Filter rows', and 'Microsoft Excel output'. Below the diagram, the 'Execution Results' window is open, showing the first 18 rows of data from the transformation. The data includes columns for item names like 'Pasar' and various numerical values ranging from 10000 to 35000.

The screenshot shows a Microsoft Excel spreadsheet titled 'Cabai Merah Keriting Februari 2019.xls'. The data is presented in a grid format with columns labeled from A to W. The first few rows correspond to the data shown in the Kettle execution results, with many subsequent rows containing identical or very similar values, indicating a large dataset.

## 3. Cleansing Maret 2019

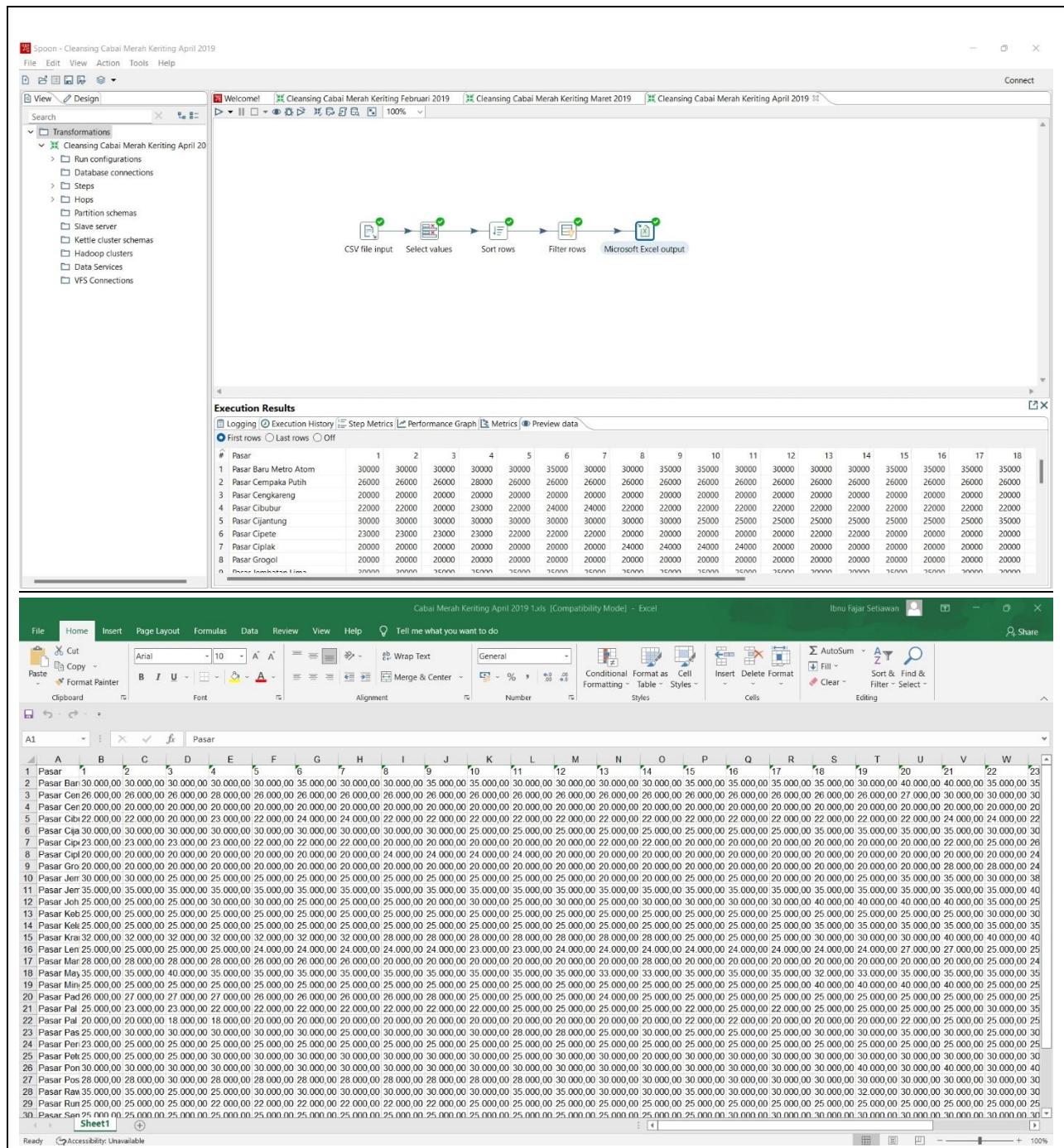


**Execution Results**

#	Pasar	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	Pasar																		
2	Pasar Baru Metro Atom	35000	35000	30000	30000	35000	30000	35000	30000	35000	30000	30000	30000	30000	30000	30000	30000	30000	30000
3	Pasar Cempaka Putih	18000	18000	18000	18000	20000	18000	20000	20000	20000	24000	24000	24000	24000	24000	24000	24000	25000	25000
4	Pasar Cengkareng	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000
5	Pasar Cibubur	22000	24000	24000	22000	22000	20000	20000	20000	20000	26000	22000	22000	22000	22000	22000	22000	22000	22000
6	Pasar Ciputat	30000	30000	30000	30000	30000	30000	30000	30000	30000	25000	25000	25000	25000	25000	25000	25000	25000	25000
7	Pasar Cipatih	18000	18000	18000	18000	18000	18000	18000	18000	18000	20000	20000	20000	20000	20000	20000	20000	20000	20000
8	Pasar Grogol	17000	17000	17000	17000	17000	17000	17000	17000	17000	17000	17000	17000	17000	17000	17000	17000	17000	17000
n	Pasar Tambang Ulo	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~

#### 4. Cleansing April 2019





5. Cleansing Mei 2019



The screenshot shows the Apache Kettle interface. On the left, the 'View' and 'Design' panes are visible, with 'Transformations' expanded to show 'Cleansing Cabai Merah Keriting Mei 201'. The main workspace displays a flow diagram with steps: CSV file input, Select values, Sort rows, Filter rows, and Microsoft Excel output. Below the workspace is the 'Execution Results' pane, which includes tabs for Logging, Execution History, Step Metrics, Performance Graph, Metrics, and Preview data. The 'Preview data' tab is selected, showing a table of data with columns 1 through 18. The data consists of 18 rows of market names and their corresponding values.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	Pasar																	
2	Pasar Baru Metro Atom	30000	40000	35000	35000	35000	35000	50000	40000	45000	40000	40000	45000	45000	40000	45000	40000	40000
3	Pasar Cempaka Putih	28000	28000	28000	32000	32000	32000	32000	32000	32000	32000	32000	32000	32000	32000	32000	32000	32000
4	Pasar Cengkareng	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	30000	30000	30000	30000	30000
5	Pasar Cibitung	22000	25000	25000	24000	24000	25000	28000	32000	30000	34000	32000	32000	32000	32000	32000	28000	28000
6	Pasar Cipete	25000	25000	25000	26000	25000	35000	35000	35000	35000	35000	35000	30000	30000	26000	26000	23000	25000
7	Pasar Cipatak	24000	24000	32000	30000	30000	30000	24000	24000	24000	24000	24000	24000	28000	28000	24000	24000	24000
8	Pasar Grogol	24000	28000	24000	36000	36000	36000	36000	36000	36000	36000	36000	32000	32000	28000	28000	28000	28000
n	Open last modified files	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	

The bottom part of the screenshot shows a Microsoft Excel window titled 'Cabai Merah Keriting Mei 2019.xls [Compatibility Mode] - Excel'. The spreadsheet contains the same data as the preview in the Spoon interface, with columns A through W and rows 1 through 30. The data is presented in a grid format with various numerical values.

## 6. Cleansing Juni 2019



The screenshot shows the Apache Kettle Transformation Editor. On the left, the 'Transformations' tree view is open, showing a single transformation named 'Cleansing Cabai Merah Keriting Juni 2019'. The main workspace displays a flow diagram with the following steps: CSV file input → Select values → Sort rows → Filter rows → Microsoft Excel output. Below the workspace is the 'Execution Results' panel, which contains a table of data. The table has 17 columns and approximately 30 rows, starting with columns labeled 1 through 17. The data includes various market names like Pasar Induk Kramat Jati, Pasar Senen Blok III - VI, Pasar Jembatan Merah, etc. The bottom part of the screenshot shows a Microsoft Excel window titled 'Cabai Merah Keriting Juni 2019.xls [Compatibility Mode] - Excel' containing the same data.

## 7. Cleansing Juli 2019



The screenshot displays two windows. The top window is the Apache Kettle Spoon interface, showing a data transformation process named 'Cleansing Cabai Merah Keriting Juli 2019'. The process flow is: CSV file input → Select values → Sort rows → Filter rows → Microsoft Excel output. The bottom window is an Excel spreadsheet titled 'Cabai Merah Keriting Juli 2019.xls' (Compatibility Mode). Both windows show the same data, which is a list of 18 entries, each starting with 'Pasar' followed by a series of numbers. The Excel window includes a standard ribbon menu and toolbars.

#	Pasar	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	Pasar Baru Metro Atom	60000	70000	70000	70000	75000	70000	70000	65000	70000	70000	70000	70000	70000	70000	70000	75000	75000	75000
2	Pasar Cempaka Putih	50000	50000	50000	50000	50000	50000	50000	60000	60000	60000	60000	60000	60000	60000	60000	60000	70000	70000
3	Pasar Cengkareng	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000
4	Pasar Cibubur	50000	55000	28000	45000	55000	55000	55000	65000	65000	65000	65000	65000	65000	65000	65000	65000	65000	60000
5	Pasar Cijantung	65000	65000	65000	70000	70000	70000	70000	70000	70000	70000	70000	70000	70000	70000	68000	70000	70000	70000
6	Pasar Cipete	55000	55000	60000	60000	60000	60000	65000	65000	65000	65000	65000	65000	65000	65000	65000	65000	65000	65000
7	Pasar Cipatak	50000	50000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000
8	Pasar Grogol	50000	50000	60000	65000	65000	65000	65000	60000	60000	65000	65000	60000	60000	60000	60000	60000	60000	70000
n	Pasar Tambaklises	€0000	€0000	€0000	€0000	€0000	€0000	€0000	€0000	€0000	€0000	€0000	€0000	€0000	€0000	€0000	€0000	€0000	€0000

## 8. Cleansing Agustus 2019



The screenshot shows the Apache NiFi interface with a transformation named "Cleansing Cabai Merah Keriting Agustus 2019". The flow consists of five steps: CSV file input, Select values, Sort rows, Filter rows, and Microsoft Excel output. The execution results window displays log messages indicating the start and completion of the transformation.

**Execution Results Log:**

```

2022/06/13 16:12:54 - Spoon - Transformation opened.
2022/06/13 16:12:54 - Spoon - Started the transformation execution.
2022/06/13 16:12:54 - Cleansing Cabai Merah Keriting Agustus 2019 - Dispatching started for transformation [Cleansing Cabai Merah Keriting Agustus 2019].
2022/06/13 16:12:54 - CSV file input[0] - Header row skipped in file:///D:/file20kuliah%20mstr4/fileprakdatawarehouse/prak 10/Cabai Merah Keriting Agustus 2019.csv'.
2022/06/13 16:12:54 - CSV file input[0] - Finished processing (I=49, O=0, R=48, W=48, U=0, E=0)
2022/06/13 16:12:55 - Select values[0] - Finished processing (I=0, O=0, R=48, W=48, U=0, E=0)
2022/06/13 16:12:55 - Sort rows[0] - Finished processing (I=0, O=0, R=48, W=48, U=0, E=0)
2022/06/13 16:12:55 - Filter rows[0] - Finished processing (I=0, O=0, R=48, W=34, U=0, E=0)
2022/06/13 16:12:55 - Microsoft Excel output[0] - Finished processing (I=0, O=34, R=34, W=34, U=0, E=0)
2022/06/13 16:12:55 - Spoon - The transformation has finished!

```

**Microsoft Excel Output:**

The screenshot shows a Microsoft Excel spreadsheet titled "Cabai Merah Keriting Agustus 2019.xls" with a single sheet named "Sheet1". The data consists of approximately 30 rows of text, each containing a value for columns A through W. The data starts with "Pasar" and ends with "Pasar Sen".

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	
1	Pasar	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
2	Pasar Bar	80.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	65.000,00	60.000,00	60.000,00	65.000,00	
3	Pasar Cen	60.000,00	60.000,00	60.000,00	72.000,00	72.000,00	72.000,00	72.000,00	72.000,00	72.000,00	72.000,00	72.000,00	72.000,00	72.000,00	72.000,00	72.000,00	72.000,00	72.000,00	72.000,00	80.000,00	80.000,00	80.000,00	80.000,00	
4	Pasar Cen	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	75.000,00	75.000,00	
5	Pasar Cibi	70.000,00	85.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	70.000,00	
6	Pasar Cijo	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	80.000,00	80.000,00	75.000,00	75.000,00	
7	Pasar Cip	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	85.000,00	85.000,00	85.000,00	85.000,00	
8	Pasar Cip	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	65.000,00	
9	Pasar Gro	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	80.000,00	80.000,00	80.000,00	80.000,00	
10	Pasar Jerr	55.000,00	55.000,00	55.000,00	55.000,00	55.000,00	55.000,00	55.000,00	55.000,00	55.000,00	55.000,00	55.000,00	55.000,00	55.000,00	55.000,00	55.000,00	55.000,00	55.000,00	55.000,00	60.000,00	60.000,00	60.000,00	60.000,00	
11	Pasar Jerr	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	90.000,00	
12	Pasar Jerr	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	85.000,00	95.000,00	95.000,00	95.000,00	
13	Pasar Keb	55.000,00	58.000,00	60.000,00	60.000,00	60.000,00	58.000,00	60.000,00	58.000,00	60.000,00	58.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	80.000,00	80.000,00	80.000,00	80.000,00	
14	Pasar Kel	70.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	80.000,00	80.000,00	80.000,00	80.000,00	
15	Pasar Krai	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	90.000,00	90.000,00	90.000,00	90.000,00	
16	Pasar Len	70.000,00	85.000,00	65.000,00	70.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	80.000,00	80.000,00	80.000,00	80.000,00	
17	Pasar Mar	70.000,00	70.000,00	70.000,00	70.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	80.000,00	80.000,00	80.000,00	80.000,00	
18	Pasar May	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	85.000,00	85.000,00	85.000,00	85.000,00	
19	Pasar Min	77.500,00	70.000,00	60.000,00	60.000,00	60.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	80.000,00	80.000,00	80.000,00	80.000,00	
20	Pasar Pad	60.000,00	75.000,00	70.000,00	70.000,00	70.000,00	75.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	80.000,00	
21	Pasar Pelt	75.000,00	70.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	75.000,00	85.000,00	85.000,00	85.000,00	85.000,00	
22	Pasar Pas	55.000,00	60.000,00	60.000,00	60.000,00	65.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	80.000,00	80.000,00	80.000,00	80.000,00	
23	Pasar Pas	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	
24	Pasar Per	70.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	
25	Pasar Pelt	70.000,00	65.000,00	65.000,00	62.000,00	70.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	70.000,00	70.000,00	70.000,00	70.000,00	
26	Pasar Pon	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	80.000,00	80.000,00	80.000,00	80.000,00	
27	Pasar Pos	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	
28	Pasar Raw	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	60.000,00	80.000,00	80.000,00	80.000,00	80.000,00	
29	Pasar Run	65.000,00	65.000,00	65.000,00	65.000,00	65.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	80.000,00	80.000,00	80.000,00	80.000,00	
30	Pasar Sen	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	
31	Pasar Tiong	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	70.000,00	

The screenshot shows the Apache Kettle Spoon interface. The top part displays a data transformation workflow: CSV file input → Select values → Sort rows → Filter rows → Microsoft Excel output. The left sidebar shows a tree view of transformations, including 'Cleansing Cabai Merah Keriting Septem'. The bottom part shows the 'Execution Results' window with a table of data. The table has columns labeled 1 through 18. The first few rows of data are:

#	Pasar	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	Pasar Baru Metro Atom	70000	70000	65000	65000	65000	65000	60000	60000	60000	60000	60000	60000	60000	65000	65000	50000	60000	60000
2	Pasar Cempaka Putih	70000	70000	70000	70000	70000	70000	70000	70000	70000	70000	70000	70000	70000	70000	70000	70000	70000	70000
3	Pasar Gengkareng	75000	75000	70000	70000	70000	70000	70000	70000	70000	70000	70000	70000	70000	70000	70000	70000	70000	70000
4	Pasar Cibubur	65000	65000	65000	60000	55000	50000	50000	65000	60000	60000	55000	60000	60000	60000	60000	50000	55000	50000
5	Pasar Cijantung	70000	70000	70000	70000	70000	70000	70000	70000	70000	70000	70000	70000	70000	70000	55000	55000	55000	55000
6	Pasar Cipete	70000	70000	70000	65000	60000	60000	60000	60000	60000	60000	55000	55000	55000	55000	55000	50000	55000	50000
7	Pasar Cipatak	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	60000	50000	50000	50000	50000	50000
8	Pasar Grogol	60000	60000	60000	60000	56000	56000	60000	50000	50000	45000	45000	45000	45000	45000	45000	45000	45000	45000

The bottom part shows a screenshot of Microsoft Excel with the title 'Cabai Merah Keriting September 2019.xls [Compatibility Mode]'. The spreadsheet contains the same data as the execution results table.

## 10. Cleansing Oktober 2019



The screenshot shows the Apache Kettle interface for data cleansing. On the left, the 'Transformations' tree view is expanded to show 'Cleansing Cabai Merah Keriting Oktober'. The main workspace displays a data flow diagram with five steps: 'CSV file input', 'Select values', 'Sort rows', 'Filter rows', and 'Microsoft Excel output'. Below the diagram, the 'Execution Results' window is open, showing the first 18 rows of data. The data consists of 18 columns, with the first column labeled '# Pasar' and subsequent columns labeled 1 through 18. The data includes entries like 'Pasar Baru Metro Atom' and 'Pasar Cempaka Putih'. At the bottom, a Microsoft Excel spreadsheet titled 'Cabai Merah Keriting Oktober 2019.xls' is shown, containing the same data structure.

## 11. Cleansing November 2019



The Transformation Editor interface shows a workflow:

```

graph LR
    A[CSV file input] --> B[Select values]
    B --> C[Sort rows]
    C --> D[Filter rows]
    D --> E[Microsoft Excel output]
  
```

**Execution Results:**

#	Pasar	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	Pasar Baru Metro Atom	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	
2	Pasar Cempaka Putih	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	45000	
3	Pasar Cengkareng	70000	70000	70000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	
4	Pasar Cibubur	40000	40000	40000	30000	32000	35000	30000	35000	35000	35000	35000	30000	30000	30000	30000	30000	40000	
5	Pasar Cijantung	40000	40000	35000	35000	35000	35000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	
6	Pasar Cipete	30000	35000	35000	30000	30000	35000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	45000	
7	Pasar Ciplak	40000	40000	40000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	35000	35000	
8	Pasar Grogol	30000	30000	30000	28000	28000	30000	30000	40000	40000	40000	32000	32000	32000	35000	35000	35000	32000	
n	Pasar Tambahan	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	

**Cabai Merah Keriting November 2019.xls [Compatibility Mode] - Excel**

The Excel spreadsheet contains the same 18 rows of data as the execution results, with columns labeled A through W.

## 12. Cleansing Desember 2019



#	Pasar	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	Pasar Baru Metro Atom	40000	40000	40000	40000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	40000	40000
2	Pasar Cempaka Putih	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	40000	40000	40000
3	Pasar Cengkareng	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000
4	Pasar Cibubur	35000	35000	36000	34000	32000	34000	35000	35000	30000	32000	34000	34000	40000	40000	40000	35000	35000	38000
5	Pasar Cijantung	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	40000	45000
6	Pasar Cipete	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000
7	Pasar Cipatak	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000	35000
8	Pasar Grogol	30000	30000	30000	28000	30000	28000	28000	28000	28000	28000	28000	28000	28000	28000	36000	36000	36000	36000
n	Pasar Tambang Timur	27000	27000	27000	27000	27000	27000	27000	27000	27000	27000	27000	27000	27000	27000	27000	27000	27000	27000

#### 4. File Praktikum

Github Repository:

<https://github.com/IbnuFajar7/Data-Warehouse/tree/main/Prak-10>



## 5. Soal Latihan

Soal:

1. Apa itu Cleansing Data?
2. Mengapa Data perlu dilakukan Cleansing?

Jawaban:

1. Data cleansing adalah suatu proses mendeteksi dan memperbaiki (atau menghapus) data set, tabel, dan database yang korup atau tidak akurat.
2. Menghilangkan kesalahan dan inkonsistensi yang muncul saat beberapa data sources dikumpulkan dalam satu dataset. Meningkatkan efisiensi kerja karena proses ini akan memudahkan Anda dan tim pengolah data untuk menemukan apa yang dibutuhkan dari data.

## 6. Kesimpulan

- a. Dalam pelaksanaan praktikum Data Warehouse, kita harus benar-benar teliti dalam menginputkan suatu fungsi untuk menampilkan suatu keluaran pada layar dengan sesuai.
- b. Kita dapat mengetahui bahwa kita bisa melakukan cleansing data yang bisa memudahkan kita untuk cek data yang kemungkinan kita salah input, dengan metode ini kita akan bisa lebih mudah tanpa harus cek satu-satu dan memakan waktu yang lebih lama, waktu ini akan membuat kita lebih efisien.

## 7. Cek List (✓)

No	Elemen Kompetensi	Penyelesaian	
		Selesai	Tidak Selesai
1.	Tugas Akhir Praktikum II	✓	

## 8. Formulir Umpam Balik

No	Elemen Kompetensi	Waktu Penggerjaan	Kriteria
1.	Tugas Akhir Praktikum II	30 Menit	1

Keterangan:

1. Menarik
2. Baik
3. Cukup
4. Kurang

